

**EXHIBIT B**

**UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK**

Daily News, LP; The Chicago Tribune Company, LLC; Orlando Sentinel Communications Company, LLC; Sun-Sentinel Company, LLC; San Jose Mercury-News, LLC; DP Media Network, LLC; ORB Publishing, LLC; and Northwest Publications, LLC,

Plaintiffs,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL, LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. 1:24-cv-03285-SHS

**PLAINTIFFS' FIRST SET OF REQUESTS FOR PRODUCTION TO  
MICROSOFT CORPORATION**

Pursuant to the Federal Rules of Civil Procedure and the Local Rules of the Southern District of New York, (the “Local Civil Rules”), Plaintiffs Daily News, LP (the “New York Daily News”); The Chicago Tribune Company, LLC, (the “Chicago Tribune”); Orlando Sentinel Communications Company, LLC (the “Orlando Sentinel”); Sun-Sentinel Company, LLC (the “Sun-Sentinel”); San Jose Mercury-News, LLC (the “Mercury News”); DP Media Network, LLC (the “Denver Post”); ORB Publishing, LLC (the “Orange County Register”); and Northwest Publications, LLC (the “Pioneer Press”) (collectively the “Publishers”) request that Microsoft Corporation (“Microsoft”) produce the requested documents at the offices of Rothwell, Figg, Ernst and Manbeck, P.C., 901 New York Ave NW, Suite 900, Washington, DC, 20001, within thirty days after the date of service of this document, or at such other mutually agreeable time and place.

## **I. DEFINITIONS**

The following definitions and instructions apply to the requests listed below:

1. All definitions and rules of construction set forth in Rule 26.3 of the Local Rules of the Southern District of New York are incorporated by reference herein.

2. “Affiliate” means, with respect to any Person, any other Person that, directly or indirectly through one or more intermediaries, controls, is controlled by, or is under any common control with, such Person.

3. “Agreement(s)” means any contract, agreement, arrangement, or understanding, formal or informal, oral or written, between or among two or more Persons.

4. “AI Model(s)” means a component of an information system and/or large language model that implements artificial intelligence (“AI”) technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.

5. “Generative AI Models(s)” means Defendants’ AI Models that emulate the structure and characteristics of input data in order to generate derived synthetic content, such as images, videos, audio, text, and other digital content.

6. “Text Generation AI Model(s)” means Defendants’ Generative AI Models, including generative pre-trained transformers or large language models (LLMs), that provide text outputs in response to inputs or prompts, including but not limited to GPT base, GPT-1, GPT-2, GPT-3, GPT 3.5, GPT-4, GPT-4 Turbo, GPT-4o, MAI-1, and all other versions or other Generative AI Models included in Defendants’ Generative AI Products and Services.

7. “Complaint” means the operative complaint in the Action as of the date these Requests are served.

8. “Defendants” means Microsoft and OpenAI.

9. “Electronically Stored Information” or “ESI” refers to information and Documents within the full scope of Federal Rule of Civil Procedure 34 with all Metadata intact—created,

manipulated, communicated, stored, and best utilized in digital form, and stored on electronic media.

10. “Generative AI Product(s) and Service(s)” means any publicly available or commercial product or service offering that includes a Generative AI Model, including but not limited to ChatGPT branded products and services, Copilot branded products and services, ChatGPT, Browse with Bing, Bing Chat, OpenAI’s soon-to-be announced search product,<sup>1</sup> APIs, and platforms otherwise providing access to a Generative AI Model.

11. “Including” means including but not limited to.

12. “Journalism,” for the purposes of these requests, means the activity of writing or creating content for newspapers, magazines, news websites, mobile applications, television, podcasts, or any other publication and/or news outlet, and includes the work of each of The Publishers as alleged in the Complaint.

13. “Metadata” refers to information about an electronic file that is embedded in the file, describing the characteristics, origins, usage, and validity of the electronic file, and includes but is not limited to descriptive, structural, administrative, and relationship information.

14. “OpenAI” means Defendants OpenAI Inc., OpenAI LP, OpenAI GP, LLC, OpenAI Opco, LLC, OpenAI Global, LLC, OpenAI Holdings LLC, OpenAI, LLC, or OAI Corporation, LLC, their subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and their respective officers, directors, employees, consultants, representatives, and agents.

15. “Relating to” means discussing, describing, referring to, pertaining to, containing, analyzing, studying, reporting on, commenting on, evidencing, constituting, setting forth,

---

<sup>1</sup> See Anna Tong, OpenAI Plans to Announce Google Search Competitor on Monday, Sources Say, REUTERS (May 10, 2024), <https://www.reuters.com/technology/openai-plans-announce-google-search-competitor-monday-sources-say-2024-05-09/>

considering, recommending, concerning, or pertaining to, in whole or in part.

16. “Microsoft” or “You” or “Your” means Defendant Microsoft Corporation, its subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and their respective officers, directors, employees, consultants, representatives, and agents.

17. “Publisher” or “Publishers” means any one or more of Plaintiffs The New York Daily News, The Chicago Tribune, the Orlando Sentinel, the Sun-Sentinel, the Mercury News, the Denver Post, the Orange County Register, the Pioneer Press, and their subsidiaries, affiliates, divisions, predecessors in interest, successors, or assignees, and respective officers, directors, employees, consultants, representatives, and agents.

18. “Publisher’s Content” or “Publishers’ Content” means material published by any one or more of the Publishers, including in print, online, audio, or in any other format, whether on Publisher-owned platforms or available through third-party sources.

19. “Training Dataset(s)” means the data and content used to pre-train, train, fine-tune, or otherwise affect the performance of the Generative AI Models, including any data and content used for reinforcement learning from human feedback or reinforcement learning from AI feedback.

## **II. INSTRUCTIONS**

1. Produce all responsive documents prepared, sent, or received, in whole or in part, during the time period applicable to the request. You should identify in some reasonable manner the person from whose files each document you produce was found.

2. You are to produce all documents, as defined above, that are in the possession, custody or control of you, or in the possession, custody or control of any attorney or agent for you. Without limiting the terms, “possession, custody, or control” of an item means that you either have physical possession of the item or have a right to possession of the item that is equal or superior to the person who has physical possession of the item.

3. You are under a duty to supplement your responses to these discovery requests

pursuant to FRCP 26. If after you produce any of the discovery requested below you discover or receive additional documents or information responsive to the requests, you should promptly produce such additional documents or information to the full extent required by the rules.

4. Pursuant to FRCP 26, you are to provide a privilege log for any documents you have withheld on the basis of privilege(s). Request for such a log is hereby made, effective as of the response date of these requests. If you assert a claim of privilege with respect to any document request, or portion thereof, you should produce a log, accounting for each document withheld individually, and providing the nature of the privilege being claimed; the type of document, date, author, addressees or recipients, and general subject matter of the document, and such other information as is sufficient to identify the document. If a claim of privilege applies to only a portion of a document, the document should be produced, and the portion claimed to be privileged obscured and stamped “redacted.”

5. The documents requested below should be produced in the manner that they are kept in the usual course of business. Documents should be produced with the label or labels from any file folder, binder, file drawer, file box, notebook, computer disk, or other container in which the document was found. Documents on computers or computer disks, including but not limited to electronic mail and corresponding attachments, should be produced in both native format and single-page TIFF format with optical character recognition (OCR). Even if only part of a document is responsive to a request, you should still produce the entire document including all attachments.

6. All duplicates or copies of documents are to be provided to the extent they have handwriting, additions, or deletions of any kind different from the original document being produced.

7. All documents which cannot be legibly copied should be produced in their original form.

8. If any documents requested herein have been lost, discarded or destroyed, they shall be identified as completely as possible, including, without limitation, the following information:

(a) the date of disposal; (b) the manner of disposal; (c) the reason for disposal; (d) the person authorizing disposal; and (e) the person disposing of the document.

9. If no Documents responsive to a particular request exist, state that no responsive Documents exist.

### **III. DOCUMENT REQUESTS**

#### **REQUEST FOR PRODUCTION NO. 1:**

Documents concerning the use of each Publisher's Content for training or fine-tuning Your Text Generation AI Models, including, for every Text Generation AI Model that uses any Publisher's Content: (i) the amount and prevalence of each Publisher's Content in the Training Datasets by token count, word count, article count, as a percentage of the Training Datasets, or by any other metric; (ii) rules or guidance for choosing what is and is not included in the Training Datasets for the Text Generation AI Models; (iii) use of each Publisher's Content to clean or curate the Training Datasets; (iv) any weighting of particular Training Datasets for the Text Generation AI Models; and (v) use of each Publisher's Content for any post-training activity, including fine tuning.

#### **REQUEST FOR PRODUCTION NO. 2:**

Documents You have gathered to submit or have submitted to any legislative or executive agency, committee, or other governmental entity in the United States or Europe that concern or relate to the allegations in the Complaint, including any Documents related to Your intended or actual submission.

#### **REQUEST FOR PRODUCTION NO. 3:**

Documents concerning copyrighted material in the Text Generation AI Models, the Training Datasets, or any Generative AI Products and Services, including documents or communications regarding the appropriateness, value, importance, necessity, and/or concerns over or legal implications of the use of such materials.



**REQUEST FOR PRODUCTION NO. 4:**

Documents concerning claims of copyright and trademark infringement, including complaints regarding infringement, steps taken to respond to such claims, responses to such complaints, steps taken to mitigate any harm resulting from such claims, and steps taken to prevent infringement.

**REQUEST FOR PRODUCTION NO. 5:**

Documents reflecting policies, procedures or practices concerning use of intellectual property in AI Models, including the policy OpenAI referred to as “Copyright Shield” in the November 6, 2023, blog post on OpenAI’s website titled “New models and developer products announced at DevDay.”

**REQUEST FOR PRODUCTION NO. 6:**

Documents concerning alternatives to using copyrighted content to train the Text Generation AI Models without compensation including using licensing agreements or material in the public domain.

**REQUEST FOR PRODUCTION NO. 7:**

Documents concerning the impact of AI Models on Journalism and the past, current or future use of the Text Generation AI Models or any Generative AI Products and Services as a source of Journalism, substitute for Journalism, or product in competition with Journalism (including any of the Publisher’s Journalism).

**REQUEST FOR PRODUCTION NO. 8:**

Documents sufficient to show to what extent and in what form each Publisher’s Content resides in, is stored in, is copied by, or otherwise exists within the Training Datasets and/or the Text Generation AI Models.

**REQUEST FOR PRODUCTION NO. 9:**

Documents sufficient to show and describe Microsoft’s corporate organizational structure

as it relates to Microsoft's use or development of AI Models or any investment(s) related thereto.

**REQUEST FOR PRODUCTION NO. 10:**

Documents sufficient to identify Microsoft's current and former officers, directors, managers, employees and consultants with knowledge of the Text Generation AI Models, business practices regarding the Text Generation AI Models, and commercialization and use of the Text Generation AI Models.

**REQUEST FOR PRODUCTION NO. 11:**

Documents concerning Defendants' use of Retrieval Augmented Generation ("RAG") in Generative AI Products and Services, including: (i) use of each of the Publisher's Content for RAG; (ii) rules or guidance for preferring or avoiding certain web domains (including each of the Publishers' web domains or sites known to host content copied from other sources) when choosing what to include in a generative result; (iii) rules for whether and how the results from RAG search products are similar to or different from standard search results; (iv) and rules regarding paywalls.

**REQUEST FOR PRODUCTION NO. 12:**

Documents concerning the features or capabilities of Retrieval Augmented Generation.

**REQUEST FOR PRODUCTION NO. 13:**

A complete listing of every data sample in the Training Datasets used to train each of the Text Generation AI Models and the source from which each data sample was obtained.

**REQUEST FOR PRODUCTION NO. 14:**

Documents concerning any commercial or supercomputing collaborations between and among the Defendants relating to the Generative AI Models, including licensing, developing, training, hosting, engineering, testing, deploying, or operating Generative AI Product(s) and Service(s).

**REQUEST FOR PRODUCTION NO. 15:**

Documents concerning the decision not to release Training Datasets or other information

for Defendants' Text Generation AI Model(s) after GPT-3 (including all versions thereof, such as GPT-4 Turbo and GPT-4o), including any communications regarding this decision and concerns regarding the effects of releasing this data on safety, ethics, profitability, competition, and legal obligations and/or ramifications.

**REQUEST FOR PRODUCTION NO. 16:**

Documents concerning the compilation, acquisition, and curation of Training Datasets, including: (i) communications regarding paying for or obtaining permission to use any Training Datasets or content therein, (ii) communications with Training Dataset creators, including Common Crawl, and (iii) communications concerning the relevance and impact of paywalls on building Training Datasets or gathering training data.

**REQUEST FOR PRODUCTION NO. 17:**

Documents concerning the removal of any content, including information regarding authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use of works, from Training Datasets before, during, or after training of Defendants' Generative AI Model(s).

**REQUEST FOR PRODUCTION NO. 18:**

Documents concerning all processing performed on the Training Datasets containing the Publishers' Content or Journalism Content used to train Defendants' Generative AI Products and Services, including: (i) the ordering of training samples containing the Publishers' or Journalism Content in the training process as compared to training samples not containing the Publishers' or Journalism Content; (ii) placement of training samples containing the Publishers' or Journalism Content in the training process; (iii) rules for choosing the order of training samples containing the Publishers' or Journalism Content in the training process; (iv) rules and procedures for training samples containing the Publishers' or Journalism Content that are not used on all other training samples; (v) the processing of the Publishers' or Journalism Content in the fine-tuning/instruction

tuning phase of training; (vi) use of the Publishers' or Journalism Content to train or inform certain, modules, layers, mixtures of experts, or other portions of the Text Generation AI Models more than others, (vii) any measures taken to include or exclude certain portions of training samples containing the Publishers' or Journalism Content during training, and (viii) the importance of training on accurate and/or well-written content, including the Publishers' or Journalism Content.

**REQUEST FOR PRODUCTION NO. 19:**

Documents sufficient to show the total number of records and tokens in Defendants' Training Datasets and the total number of records and tokens that contain (i) the Publishers' Content, (ii) Journalism content, and (iii) copyrighted content.

**REQUEST FOR PRODUCTION NO. 20:**

Documents sufficient to show each source of records in Defendants' Training Datasets, and for each source: (i) the total number of tokens, (ii) the total number of records, (iii) the total number of tokens that contain the Publishers' Content, (iv) the total number of records that contain the Publishers' Content, (v) the total number of tokens that contain Journalism content, (vi) the total number of records that contain Journalism content; (vii) the total number of tokens that contain copyrighted content, (viii) the total number of records that contain copyrighted content and (ix) any agreements relating to Defendants' use, licensing, or purchasing of the source of records.

**REQUEST FOR PRODUCTION NO. 21:**

Documents sufficient to identify each article of the Publishers' Content contained in Defendants' Training Datasets.

**REQUEST FOR PRODUCTION NO. 22:**

Documents concerning all steps taken or contemplated to train, instruct, program, or refine Defendants' Generative AI Products and Services using the Publishers' Content or Journalism content, including: (i) the use of the Publishers' or Journalism Content in instruction/fine tuning; (ii) use of the Publishers' Content or Journalism content in Reinforcement Learning from Human

Feedback (“RLHF”) or Reinforcement Learning from AI Feedback (“RLAIF”), (iii) use of the Publishers’ or Journalism Content to train the reward component or function of RLHF or RLAIF, (iv) use of the Publishers’ or Journalism Content to evaluate a feature of a Training Dataset, such as classifying Journalism content or identifying quality of a training sample, and (v) RLHF moderator and ranking guidance, training materials and ranking factors, weighting or scoring factors, and results/output of the human annotators, moderators and reviewers.

**REQUEST FOR PRODUCTION NO. 23:**

Documents concerning any analyses, studies, measurements, testing, experimentation, assessments, or other evaluation of the incremental value of including Journalism content, including the Publishers’ Content, in Defendants’ Training Datasets.

**REQUEST FOR PRODUCTION NO. 24:**

Documents concerning any analyses, studies, measurements, testing, experimentation, assessments, or other evaluation of the quantity, ordering, placement, or other configurable parameter of Journalism content, including the Publishers’ Content, in Defendants’ Training Datasets.

**REQUEST FOR PRODUCTION NO. 25:**

Documents concerning Your use or knowledge of web crawlers, bots, spiders, user agents, and related tools to access the Publishers’ Content, including the Publishers’ websites and digital products, including documents sufficient to show: (i) the time periods in which those tools accessed the Publishers’ Content, websites and digital products, (ii) the Publishers’ Content those tools accessed, fetched, or parsed from the Publishers’ websites and digital products, and (iii) the Publishers’ Content from these URLs that was crawled, processed, copied, stored, and/or indexed by You and/or systems under Your control.

**REQUEST FOR PRODUCTION NO. 26:**

Documents concerning Your reliance on, use of, or understanding of third-party robots.txt

crawler directives or third-party meta tag crawler directives in managing the crawling and indexing activity of your web crawlers, bots, spiders, user agents, GPTBot, BingBot, any plugins or custom GPTs included on Defendants' platforms, and related applications to access, crawl, and/or index the Publishers' Content or third-party content.

**REQUEST FOR PRODUCTION NO. 27:**

Documents concerning all measures taken to conceal or obscure Defendants' use of the Publishers' Content during any stage of the development of each of Defendants' Generative AI Products and Services, including: (i) any methods, procedures, or instructions used to prevent Defendants' Generative AI Products and Services from reproducing Training Datasets containing the Publishers' Content, (ii) any methods, procedures, or instructions used to determine if the output of Defendants' Generative AI Products and Services contains the Publishers' Content, (iii) any methods, procedures, or instructions to avoid overfitting or memorization of the Publishers' Content in Defendants' Generative AI Models, (iv) any differences in the foregoing among Defendants' Generative AI Products and Services, and (v) any changes in the foregoing over time.

**REQUEST FOR PRODUCTION NO. 28:**

Documents concerning Defendants' reliance on earlier versions of Defendants' Generative AI Models to be used in or to train or develop later versions of Defendants' Generative AI Models, including Defendants' use of outputs, weights, or any learnings acquired from earlier versions of Defendants' Generative AI Models.

**REQUEST FOR PRODUCTION NO. 29:**

The full model cards, data set cards, or system cards for each of Defendants' Generative AI Models, including those not released to the public.

**REQUEST FOR PRODUCTION NO. 30:**

Documents relating to Your technical documentation for Defendants' Generative AI

Models, Products, and Services as defined in Annex IV of the European Union's Artificial Intelligence Act, including correspondence regarding the Act and drafts of documentation for the Act.

**REQUEST FOR PRODUCTION NO. 31:**

Documents concerning the training curricula, stages, iterations, cycles, batches, or epochs for Defendants' Generative AI Models.

**REQUEST FOR PRODUCTION NO. 32:**

Documents concerning Defendants' decisions regarding which works to use to train and develop Defendants' Generative AI Models, and how to use those works in training, including (i) documents concerning the factors considered in determining the weighting or frequency at which works are seen during the training of Defendants' models, (ii) documents concerning de-duplication of Training Datasets; (iii) documents concerning any classifiers and/or filters applied to Training Datasets that determine document quality and/or remove low quality documents; (iv) documents concerning rules and procedures for determining sample frequency; (v) documents sufficient to show how the training weights or parameters within Defendants' Generative AI Models would change if the Publishers' Content were not used during training; and (vi) documents sufficient to show which individuals made or make decisions regarding which works to use to train and develop Defendants' Generative AI Models, and how to use those works in training.

**REQUEST FOR PRODUCTION NO. 33:**

Documents sufficient to describe each aspect or phase of the process for training Defendants' Generative AI Models, including the pre-training, training, fine-tuning, and any other refinement or development processes as well as any distinctions or differences between and among the aspects or phases.

**REQUEST FOR PRODUCTION NO. 34:**

Documents concerning any changes made to Defendants' Generative AI Models or

Generative AI Products and Services from 2021 to present, including any logs showing changes, documents relating to ablation or “unlearning,” (i.e. the ability to remove memorization, content, or other effects of specific parts of the Training Datasets from the models), and documents and communications concerning any changes to the models made or considered, including in response to claims of copyright infringement or improper use of data.

**REQUEST FOR PRODUCTION NO. 35:**

Documents concerning the differences between Defendants’ Generative AI Models, including (i) differences in the Training Datasets and/or the composition of the Training Datasets used for each model; (ii) differences in the training curricula used for each model; and (iii) the differences between the “Turbo” and non-Turbo versions of Defendants’ models.

**REQUEST FOR PRODUCTION NO. 36:**

The following documents concerning each of Defendants’ Generative AI Models, Products, and Services: (i) product briefs, (ii) readme files, (iii) product requirement documents, (iv) product or project management tools and records (e.g., Jira), (v) documentation from workflow management programs, (vi) minutes or notes from technical team meetings, (vii) documentation concerning data loaders, (viii) documents related to red-teaming and/or stress testing, (ix) technical system or architecture documents, specifications, or reports, and (x) launch documents.

**REQUEST FOR PRODUCTION NO. 37:**

Source code related to (i) Defendants’ Generative AI Models, Products, and Services, (ii) training curricula for Defendants’ Generative AI Models, (ii) any ranking, weighting, or scoring of training data, (iii) data loaders, (iv) Your web crawlers, spiders, and bots, including GPTBot, BingBot and (v) any processing of training data.

**REQUEST FOR PRODUCTION NO. 38:**

Documents concerning the study, testing, or evaluation of Defendants’ Generative AI



Products and Services and Generative AI Models for hallucinations, memorization, and any problematic behaviors, including: (i) the results of the efforts of safety teams, including any red-teaming efforts, (ii) responses to ChatGPT's feedback contest, (iii) submissions to Your bug bounty program, (iv) any other red-teaming, (v) source code related to Your safety efforts, and (vi) the extent to which rates of hallucination and memorization have changed over time within Defendants' Generative AI Models.

**REQUEST FOR PRODUCTION NO. 39:**

Documents concerning the prevalence of outputs that are verbatim or substantially similar copies of text included in Training Datasets, including any documents concerning (i) steps taken to prevent or encourage outputs that are verbatim copies of text included in training data, (ii) the use of filter detectors, refusal messages, and other safety features used by Defendants to prevent unlawful access to copyrighted content contained within or referenced by Defendants' Generative AI Models, (iii) documents concerning the rate at which any such safety features are triggered in response to user queries, and (iv) documents concerning Your decisions whether to terminate user accounts for copyright abuse.

**REQUEST FOR PRODUCTION NO. 40:**

Documents or surveys related to users' trust in the accuracy or veracity of information contained in outputs of Defendants' Generative AI Products and Services and how users respond to disclaimers regarding the accuracy or veracity of information contained in such outputs.

**REQUEST FOR PRODUCTION NO. 41:**

Documents concerning Defendants' efforts and abilities to monitor, log, analyze, quantify, assess, evaluate, or review user queries within any Generative AI Products and Services.

**REQUEST FOR PRODUCTION NO. 42:**

Any surveys, studies, analyses, or statistics on patterns and practices any of Defendants' Generative AI Products and Services, including any mentions of the Publishers', such surveys,

studies, analyses, or statistics regarding of user engagement within user queries and/or inclusion of links to the Publishers' Content in user queries or outputs.

**REQUEST FOR PRODUCTION NO. 43:**

Documents concerning any attempts to use Generative AI Products and Services to circumvent paywalls or other access control measures.

**REQUEST FOR PRODUCTION NO. 44:**

Query, session, and chat logs reflecting or analyzing user sessions related to the Publishers' Content within Defendants' Generative AI Products and Services, including, for each session, user queries paired with responses to those queries.

**REQUEST FOR PRODUCTION NO. 45:**

Documents sufficient to show how Defendants' Generative AI Products and Services generate suggested follow-up queries for users.

**REQUEST FOR PRODUCTION NO. 46:**

Documents concerning user engagement metrics, events, and analytics data collected during each user session, including duration, the number of prompts or queries that a user inputs per session, click-through data, fallback rates, and bounce rates.

**REQUEST FOR PRODUCTION NO. 47:**

Documents sufficient to show (i) the number of daily active users of each of Defendants' Generative AI Products and Services, separated into the numbers of new users and repeat users on each day, from 2021 to present, and (ii) dates of subscription and changes in subscriber numbers over time.

**REQUEST FOR PRODUCTION NO. 48:**

Documents concerning Defendants' monitoring, tracking, and knowledge of investigations into whether Defendants' AI Models contain copyrighted content, including the Publishers' Content.

**REQUEST FOR PRODUCTION NO. 49:**

Documents concerning Defendants' abilities and efforts to track the Publishers' use of Defendants' Generative AI Products and Services.

**REQUEST FOR PRODUCTION NO. 50:**

Documents concerning the behavior of users of Defendants' Generative AI Products and Services, including (i) documents concerning the extent to which users of Defendants' Generative AI Products and Services navigate away from Defendants' Generative AI Products and Services to other websites, and (ii) documents concerning Defendants' ability and efforts to modify the frequency with which users navigate away from Defendants' Generative AI Products and Services to other websites.

**REQUEST FOR PRODUCTION NO. 51:**

Documents concerning the extent to which Defendants' Generative AI Products and Services direct users to the Publishers' Content in response to user queries, including (i) documents concerning how many times each Publisher's content was output by Defendants' Generative AI Products and Services and (ii) documents concerning the click-through rate for a URL output by Defendants' Generative AI Products and Services that is intended to direct the user to one of the Publisher's websites.

**REQUEST FOR PRODUCTION NO. 52:**

Documents and communications regarding the decision to exclude Bing Chat referral traffic data from Bing Webmaster Tools.

**REQUEST FOR PRODUCTION NO. 53:**

Documents concerning the decision to suspend Browse with Bing and whether or not to suspend Bing Chat.

**REQUEST FOR PRODUCTION NO. 54:**

Documents concerning any public statements made by You or Your employees relating to

this litigation.

**REQUEST FOR PRODUCTION NO. 55:**

Documents concerning any firings, terminations, lay-offs, impositions of mandatory leave, disciplinary actions, departures, resignations, or other personnel-related actions Defendants have taken or considered concerning individuals or teams involved with any ethics, harm reduction, safety, lawfulness, and responsible AI initiatives, including: (i) the removal and reinstatement of Sam Altman, (ii) the firing of Leopold Aschenbrenner and Pavel Izmailov, (iii) the appointment and role of Dee Templeton, (iv) the firing of Helen Toner, (v) the departures of Ilya Sutskever, Jan Leike, and Daniel Kokotajlo, (vi) the dissolution and/or firing of any member of Your AI ethics team, and (vii) Microsoft's involvement in hiring and firing employees of any OpenAI entity.

**REQUEST FOR PRODUCTION NO. 56:**

Documents concerning any discussion of the use of copyrighted content by Defendants' Generative AI Models or Defendants' Generative AI Products and Services, including documents concerning whether the use of copyrighted content by Defendants' Generative AI Models or Defendants' Generative AI Products and Services constituted "fair use" under the copyright law.

**REQUEST FOR PRODUCTION NO. 57:**

Documents concerning any deliberation or decision(s) by Defendants not to publicly release research, papers, or studies related to Defendants' Generative AI Models and Defendants' Generative AI Products and Services, including any communications regarding these decisions and concerns regarding the effects of releasing this data on safety, ethics, profitability, competition, and legal obligations and/or ramifications.

**REQUEST FOR PRODUCTION NO. 58:**

Documents related to the use of Defendants' Generative AI Products and Services, including by third parties, to promulgate misinformation, disinformation, or poor quality "pink-slime" journalism, including the use of Defendants' APIs to create websites or social media posts

containing AI-generated content.

**REQUEST FOR PRODUCTION NO. 59:**

Documents concerning Your knowledge of third-party websites that contain copies of the Publishers' Content and any use of GPTBot, BingBot and any other web crawlers, bots, spiders, user agents, and related tools to access those third-party websites.

**REQUEST FOR PRODUCTION NO. 60:**

Documents concerning any indemnification obligations, common interest agreements, or joint defense agreements between Defendants related to Defendants' Generative AI Products or Services.

**REQUEST FOR PRODUCTION NO. 61:**

Documents and communications concerning negotiations between Defendants and owners of copyrighted content, including the Publishers', over Defendants' use of copyrighted content in Defendants' Generative AI Models and Defendants' Generative AI Products or Services, including documents and communications concerning the licenses to the Associated Press, News Corp, Axel Springer, The Atlantic, Dotdash Meredith, Vox Media, Reddit, and The Financial Times,

**REQUEST FOR PRODUCTION NO. 62:**

Documents concerning Your communications with other LLM creators and developers of Generative AI Products and Services regarding the use of copyrighted content to train and operate Generative AI Models and Generative AI Products and Services.

**REQUEST FOR PRODUCTION NO. 63:**

Documents concerning competition and comparisons between Defendants' Generative AI Products and Services and those of other companies, including: (i) how Generative AI tools enable You to compete in search, (ii) how Defendants' use of copyrighted materials can help Defendants compete with Google, (iii) any analyses and projections of how increasing Your search market

share will boost Your revenue, (iv) the benefits of being a search or knowledge “destination,” and (v) Google’s reduction of search referral traffic (via SGE and direct answers) and stated goal of keeping users on its site.

**REQUEST FOR PRODUCTION NO. 64:**

Documents concerning Your public relations strategies as they pertain to Generative AI Models, Generative AI Products and Services, or copyright.

**REQUEST FOR PRODUCTION NO. 65:**

Documents concerning fundraising efforts related to Defendants’ Generative AI Products and Services, including any investor solicitation efforts and any solicitation of charitable or in-kind contributions.

**REQUEST FOR PRODUCTION NO. 66:**

Documents sufficient to show (i) projected revenue streams from Defendants’ Generative AI Products, (ii) past, present, and anticipated future subscriber numbers, and (iii) product-level revenue data from 2021 to present.

**REQUEST FOR PRODUCTION NO. 67:**

Documents concerning business development plans, strategic roadmaps, and product development plans related to (i) Defendants’ Generative AI Products and Services and (ii) any future products or services in the categories of news, local reporting, and sports.

**REQUEST FOR PRODUCTION NO. 68:**

Documents concerning Your policies or practices regarding public access to Your governing documents, financial statements, and conflict of interest rules.

**REQUEST FOR PRODUCTION NO. 69:**

Documents concerning referral, affiliate, advertising, and paid search revenue generated by Defendants’ Generative AI Products and Services.

**REQUEST FOR PRODUCTION NO. 70:**

Documents sufficient to show in detail, on a monthly, quarterly, or annual basis, the revenue and profit realized from Defendants' Generative AI Products and Services, including documents sufficient to show in detail, by product or service, the gross revenue generated, the costs, the gross profit margin, and the net profit margin.

**REQUEST FOR PRODUCTION NO. 71:**

Documents relating to the pricing of Defendants' Generative AI Products and Services in the United States, including price quotes, price lists, subscription amounts, price announcements, and actual prices paid by customers.

**REQUEST FOR PRODUCTION NO. 72:**

All documents that you intend to rely upon at trial, in support of any motion, or to present to the Court for any purpose in this litigation.

**REQUEST FOR PRODUCTION NO. 73:**

All documents produced or otherwise provided by You to The New York Times Company in the litigation related to this case.

Dated: June 14, 2024

/s/Steven Lieberman

Steven Lieberman (SL8687)  
Jennifer B. Maisel (5096995)  
Robert Parker (*admitted pro hac vice*)  
Jenny L. Colgate (*admitted pro hac vice*)  
Mark Rawls (*admitted pro hac vice*)  
Kristen J. Logan (*admitted pro hac vice*)  
Bryan B. Thompson (6004147)  
ROTHWELL, FIGG, ERNST & MANBECK, P.C.  
901 New York Avenue, N.W., Suite 900 East  
Washington, DC 20001  
Telephone: (202) 783-6040  
Facsimile: (202) 783-6031  
slieberman@rothwellfigg.com  
jmaisel@rothwellfigg.com  
rparker@rothwellfigg.com  
jcolgate@rothwellfigg.com

mrawls@rothwellfigg.com  
klogan@rothwellfigg.com  
bthompson@rothwellfigg.com

Jeffrey A. Lindenbaum (JL1971)  
ROTHWELL, FIGG, ERNST & MANBECK, P.C.  
3 Manhattanville Road, Suite 105  
Purchase, New York 10577  
Telephone: (202) 783-6040  
Facsimile: (202) 783-6031  
jlindenbaum@rothwellfigg.com

*Attorneys for Plaintiffs*



**CERTIFICATE OF SERVICE**

It is hereby certified that, on June 14, 2004, the foregoing **PLAINTIFFS' FIRST SET OF REQUESTS FOR PRODUCTION TO MICROSOFT CORPORATION** was served by email upon the following counsel of record for Microsoft Corporation:

<p><b>Faegre Drinker Biddle &amp; Reath LLP</b> 1177 Avenue of the Americas, 41st Floor New York, NY 10036 212-248-3191 <b>Jeffrey S. Jacobson</b> Email: <a href="mailto:jeffrey.jacobson@faegredrinker.com">jeffrey.jacobson@faegredrinker.com</a></p> <p><b>Faegre Drinker Biddle &amp; Reath LLP</b> 1144 Fifteenth Street Suite 3400 Denver, CO 80202 303-607-3588 <b>Jared Barrett Briant</b> Email: <a href="mailto:jared.briant@faegredrinker.com">jared.briant@faegredrinker.com</a></p>	<p><b>Orrick, Herrington &amp; Sutcliffe LLP</b> The Orrick Building 405 Howard Street San Francisco, CA 94105 (415) 773-4585 <b>Annette Louise Hurst</b> Email: <a href="mailto:ahurst@orrick.com">ahurst@orrick.com</a></p> <p><b>Orrick, Herrington &amp; Sutcliffe LLP</b> 51 West 52nd Street New York, NY 10019 212-506-3778 <b>Christopher Cariello</b> Email: <a href="mailto:ccariello@orrick.com">ccariello@orrick.com</a></p> <p><b>Orrick, Herrington &amp; Sutcliffe LLP</b> 222 Berkeley Street Suite 2000 Boston, MA 02116 617-880-1800 <b>Sheryl Koval Garko</b> Email: <a href="mailto:sgarko@orrick.com">sgarko@orrick.com</a> <b>Laura Najemy</b> Email: <a href="mailto:lnajemy@orrick.com">lnajemy@orrick.com</a></p>
---	--

*/s/Nasri V. B. Hage*  
Nasri V. B. Hage